

Panorama actual, validación de test, problemas relacionados y desarrollo futuro en la investigación básica oncológica

Juan Cruz Cigudosa

Jefe de Grupo de Citogenética Molecular
Programa de Genética del Cáncer Humano
Centro Nacional de Investigaciones Oncológicas (CNIO)

RESUMEN EJECUTIVO

En primer lugar, este capítulo introduce históricamente el papel de la Genómica en la investigación oncológica. Después presenta esquemas razonados sobre el proceso de definición de biomarcadores de origen genómico y resume los principales tipos de plataformas genómicas con sus ventajas y peculiaridades. Se analizan la situación y la disponibilidad actual de biomarcadores genómicos en la práctica clínica oncológica. Finalmente, se identifican los problemas más frecuentes en la incorporación de este tipo de análisis genómicos en la rutina de la actividad clínica y se indican las posibilidades de desarrollo en este campo.

INTRODUCCIÓN: LA GENÓMICA COMO ORIGEN DE BIOMARCADORES

El cáncer es una enfermedad celular de origen genético (1). Esta frase, sencilla y simple, tiene un impacto de enormes dimensiones sobre cualquier tipo de aproximación, científica o clínica, que las ciencias biosanitarias han de seguir para estudiar el fenómeno oncológico (2). La principal consecuencia de la definición del cáncer como enfermedad genética es que, de forma casi excluyente, la investigación oncológica debe estar basada en herramientas de análisis genético.

Si se asume el origen genético del cáncer, **¿cómo se investiga la relación gen-enfermedad?** El esquema de investigación que habitualmente se sigue para caracterizar esta relación consiste, en esencia, en estudiar un gen y su efecto en el modelo biológico estudiado, en determinar las relaciones con otros genes y, por último, en diseñar modificaciones (inhibidores o activadores) de su expresión para regular su efecto en condiciones patológicas. Este sistema ha sido muy útil, pero se adivina un tanto insuficiente si se tiene en cuenta que el número de genes es de unos 30.000, que se traducen en unas 100.000 pro-

teínas, y que las interacciones gen-gen, proteína-proteína y gen-proteína, por simple combinatoria, alcanzan números casi inimaginables.

En los últimos años las herramientas de análisis genético han sufrido una revolución técnica comparable a la incorporación del microscopio en los laboratorios. Han aparecido sistemas de análisis genómico masivo que permiten analizar en un solo experimento el estado de miles de genes. Es decir, ahora estudiar la relación gen-enfermedad no se basa en analizar un gen único y sus efectos, sino en analizar el comportamiento de miles de genes de forma simultánea. Estos sistemas, denominados genéricamente matrices, *arrays*, *microarrays* o biochips, están cambiando la forma de plantear problemas y extraer conclusiones de los experimentos, ya que ofrecen una foto compleja del conjunto del genoma. Este conjunto de datos sobre expresión génica y relaciones entre genes tratados de forma sistemática mediante herramientas de análisis masivo constituye el cuerpo esencial de lo que se denomina Genómica.

La Genómica intenta proporcionar a los investigadores en Biología el equivalente de la tabla periódica de los elementos, es decir, un inventario de todos los genes que contribuyen y se coordinan para explicar la existencia y la funcionalidad de cualquier ser vivo. Este inventario estará asociado a un sistema de clasificación que genere bloques o subconjuntos de genes que se comportan de forma coordinada.

Sin duda, la mayor contribución de los estudios con *microarrays* reside en la taxonomía de las enfermedades. Clasificar un determinado tumor con un nombre y apellidos de la forma menos ambigua posible permite ajustar el diagnóstico y el tratamiento casi de forma individual, **es la base, la piedra angular, de la medicina individualizada**. Sin embargo, como todos sabemos, realizar una filiación adecuada de un caso real depende

de muchas variables: las circunstancias de recogida de la muestra (hospital, laboratorio), el rigor en la descripción inicial de todas las condiciones (anamnesis, datos de laboratorio), la disponibilidad de herramientas de diagnóstico adecuadas y, finalmente, la experiencia del profesional. Conjugar todos estos elementos tendría que tener como final el desarrollo de una clasificación de los tumores que fuese universal y efectiva.

Es importante señalar otra vez la paralogía de la Genómica con el descubrimiento y la caracterización de biomarcadores. En primer lugar, se produce el reconocimiento de una diferencia morfológica de una manera reproducible y sistemática. En segundo lugar, se intenta determinar la relación entre las características del tumor (sus biomarcadores) y las del tejido normal del que se origina. En tercer lugar, se trata de establecer asociaciones entre la evolución del paciente o la respuesta al tratamiento y los biomarcadores propuestos, de forma que éstos tengan una utilidad clínico-terapéutica. La tecnología de *microarrays* puede usarse para derivar clasificaciones basadas en biomarcadores, establecer asociaciones con la línea celular de origen y ayudar en el pronóstico, pero con la ventaja añadida de ser capaz de identificar los genes que determinan esa clasificación (3).

En este capítulo se trabajará con el concepto de biomarcadores, sus diferentes acepciones y las relaciones que la Genómica puede establecer con todos ellos. En la **Tabla I** se recogen los diferentes conceptos de biomarcador y, como ejemplo, la adscripción y uso de un biomarcador genético (la amplificación del oncogén HER2/NEU) a la administración de una terapia individualizada para el cáncer de mama.

El efecto que está teniendo la implantación de la Genómica en el medio ambiente clínico está acelerando el proceso del descubrimiento de nuevos biomarcadores diagnósticos, pero está incrementando de forma exponencial las dificultades para que esos biomarcadores adquieran valor pronóstico o predictivo. Estas dificultades

tienen su origen en la complejidad del análisis genómico (sobre la que se hablará en otra sección de este capítulo).

OBJETIVOS

1. Introducir históricamente la aplicación de la Genómica a la medicina individualizada.
2. Describir de forma práctica las plataformas básicas de análisis genómico: funcionamiento como generadores de biomarcadores y sus limitaciones.
3. Identificar las barreras para la implantación de biomarcadores de origen genómico en la práctica clínica: innovación constante en alcance y resolución, ausencia de validación adecuada y dificultades de análisis y mantenimiento de datos genéticos.
4. Indicar las perspectivas a corto y largo plazo.

TECNOLOGÍA GENÓMICA: ORIGEN, TIPOS Y UTILIDADES

Microarrays de expresión génica

Los ensayos de hibridación en *microarrays*, descritos a finales de los años 80, se basan en la disposición de material genético sobre un sustrato (plástico, cristal, membranas) en posiciones conocidas. En la actualidad existen en el mercado científico herramientas de análisis genómico (matrices o *microarrays*) diseñadas a partir de ácidos nucleicos en forma de ADNc (ADN monocatenario y complementario al ARN mensajero que contiene la secuencia codificante de un gen), secuencias cortas de nucleótidos (oligonucleótidos) o ADN de doble cadena.

En todos los casos, cada material genético que se coloca en el *array* está referido a una región del genoma cuya situación y características son conocidas o predichas con cierta verosimilitud. Es decir, cada uno de los puntos del *microarray* puede representar un gen y la información que genera se refiere al estatus de ese gen en la muestra analizada. Los *microarrays* más extendidos y utilizados son los dos primeros tipos (ADNc y oligos), que se emplean casi

Tabla I. Esquema general del concepto de biomarcador aplicado a una alteración genética

Concepto	Descripción	Ejemplo
Biomarcador diagnóstico	Define la presencia y el tipo de enfermedad con la máxima discriminación posible	Amplificación del gen HER2/NEU
Biomarcador pronóstico	Se asocia con la evolución clínica en ausencia de terapia	Biomarcador de pronóstico negativo
Biomarcador predictivo	Se asocia con la respuesta a una determinada pauta terapéutica	Biomarcador de predicción positivo
Terapia individualizada	Se emplea la información proporcionada por todos los tipos de biomarcadores para elegir la opción más adecuada para el paciente en las condiciones de presentación	Existe una terapia individualizada para la paciente: trastuzumab

exclusivamente en estudios sobre análisis masivo o global de la expresión genética de organismos.

El material obtenido de la muestra problema, convenientemente tratado y marcado con una molécula que permite su identificación o lectura (fluorescencia), se pone en contacto con el *array*, hibridando en aquellas posiciones en que se complementa con las sondas. El patrón de hibridación es revelado mediante un escáner basado en microscopía confocal y la imagen resultante se convierte a valores numéricos, que constituyen los resultados del ensayo. En el ejemplo más frecuente de los *microarrays* de expresión, este sistema permite la cuantificación de expresión génica diferencial. El nivel de expresión de un gen se refleja en el número de copias de ARNm presentes en la muestra problema y, por tanto, es proporcional al nivel de señal detectado.

Una de las ventajas significativas es que, debido a la miniaturización del sistema, la alta concentración del material permite identificar muestras presentes en un bajo número de copias. Además, las técnicas de revelado por fluorescencia ofrecen una elevada relación señal-ruido, incluso en casos de transcritos poco abundantes.

El alto grado de integración del *array* permite, en un solo ensayo, obtener multitud de valores de expresión génica relativa para distintas condiciones biológicas, lo que convierte a esta técnica en una herramienta de alto rendimiento en el área de la Genómica Funcional. El gran volumen de datos generados debe tratarse con herramientas y métodos bioinformáticos.

Microarrays de ADN genómico

Aunque la aplicación más conocida de los biochips de ADN es la determinación de perfiles de transcripción, el formato *microarray* también se ha utilizado de manera eficaz al menos en otros dos tipos de experimentos: los *microarrays* diseñados para la detección masiva y simultánea en una muestra problema de cientos de miles de polimorfismos de un único nucleótido (*single nucleotide polymorphisms* [SNP]); y los *microarrays* de ADN genómico diseñados para el estudio de las alteraciones en el número de copias de ADN (*copy number variations* [CNV]) presentes en la muestra de estudio.

Los *microarrays* de ADN genómico diseñados para detectar CNV han surgido como la evolución en clave genómica de la técnica conocida como hibridación genómica comparada (*comparative genomic hybridization* [CGH]), desarrollada en 1992 (4). La técnica de CGH se ha aplicado extensamente durante los últimos años a los más diversos tipos de neoplasias, en cuya búsqueda de regiones cromosómicas ha mostrado su gran utilidad. La CGH está basada en la hibridación sobre cromosomas y, por lo tanto, tiene una capacidad de resolución limitada, por lo que no se detectan los cambios que impliquen segmentos inferiores a 7-10 megabases. Segmentos de ese tamaño pueden contener decenas de genes cuyo comportamiento se

escapa al análisis. Esta limitación se ha salvado mediante la introducción de los *microarrays* como plataforma para hibridar las muestras problema.

Esta aproximación se denomina *array* de CGH o *CGH array*. La hibridación se realiza sobre segmentos de ADN de menor tamaño clonados en forma de: 1) cromosomas artificiales bacterianos o plasmídicos (BAC o PAC, respectivamente) (5-7); 2) ADNc correspondientes a genes de interés (como si fueran *microarrays* de expresión) (8); o 3) oligonucleótidos de unos 60 pares de bases sintetizados de forma exclusiva para su inclusión en el *microarray* y que pueden contener variantes tipo SNP (9, 10). Estos segmentos de ADN son secuencias genómicas conocidas cuya localización en el genoma es conocida y cuyo contenido génico es más o menos conocido. En esta nueva situación el nivel de resolución pasa, por tanto, de megabases (cromosomas) a kilobases (oligonucleótidos). La técnica se alimenta de la gran cantidad de datos aportados por el Proyecto Genoma Humano sobre la ubicación, la secuencia y el contenido de genes de estos fragmentos de ADN, que son, en buena medida, conocidos.

Los *microarrays* diseñados para el genotipado masivo de SNP han sido los últimos en aparecer en el campo de la Genómica Aplicada, pero su desarrollo parece imparable. Estos *microarrays* están diseñados para detectar la presencia o ausencia de un determinado SNP en una muestra problema. Por lo tanto, permiten conocer simultáneamente la secuencia de cientos de miles (con frecuencia, más de un millón) de variantes génicas presentes en la muestra.

Esta tecnología se denomina *array* de SNP o *SNP array*. Ofrece una gran ventaja respecto a las pruebas genéticas empleadas hasta ahora para determinar el genotipo de una serie de marcadores: es una prueba global, que no selecciona previamente los genes que se estudian, sino que realiza un escaneo general del genoma en busca de genotipos de interés. Supone el paso de la Farmacogenética a la Farmacogenómica (11). Además, estos tipos de *arrays* pueden emplearse (aunque con algo menos de eficacia) para detectar CNV en una muestra, por lo que también pueden considerarse *arrays* de CGH. En la **Tabla II** se presentan las características generales y la comparación de los tipos de *microarrays* descritos.

LOS MICROARRAYS Y LA GENÓMICA COMO GENERADORES DE BIOMARCADORES TUMORALES

Esquema general de la generación de biomarcadores

La generación de un biomarcador, ya sea de tipo diagnóstico, pronóstico o predictivo, y su traslado a la práctica clínica es un proceso secuencial compuesto de, al menos, cuatro etapas que incluyen e implican a investigadores (básicos y aplicados/clínicos), reguladores y agentes de empresas de Biotecnología.

Tabla II. Resumen de las características de los tres tipos de plataformas genómicas

Tipo de <i>array</i>	Muestra	Objetivo	Ventajas	Desventajas
Expresión	ARN	Detecta el perfil de expresión de todos los genes incluidos en el <i>array</i>	<ol style="list-style-type: none"> 1. Representa una foto global de la actividad de la célula 2. Permite obtener firmas moleculares del tumor 	<ol style="list-style-type: none"> 1. Es muy dependiente de las condiciones de recogida de las muestras 2. Es muy dependiente de la validación por otras técnicas 3. No detecta polimorfismos germinales ni su posible efecto
CGH	ADN	Detecta ganancias y pérdidas de las regiones del genoma	<ol style="list-style-type: none"> 1. Explica el fenotipo en función de deleciones y amplificaciones 2. Es independiente de las condiciones del material biológico 3. No es muy dependiente de desarrollos de <i>software</i> 4. Se puede trasladar a la práctica clínica con facilidad 5. Detecta polimorfismos genéticos germinales 6. Su validación es muy accesible 	<ol style="list-style-type: none"> 1. Las variaciones genéticas del número de copias pueden no tener una repercusión clara en el fenotipo: la regulación de la expresión está en un nivel superior. Es decir, se pueden tener ganancias o pérdidas sin efecto patológico 2. No detecta las regiones con pérdidas de heterocigosidad si no van acompañadas de una deleción
SNP	ADN	Detecta el genotipo (secuencia) de los marcadores polimórficos incluidos en el <i>array</i>	<ol style="list-style-type: none"> 1. Se puede explicar el fenotipo en base 2. Es independiente de las condiciones del material biológico 3. Detecta las regiones con pérdidas de heterocigosidad 4. Detecta ganancias y pérdidas de las regiones del genoma 	<ol style="list-style-type: none"> 1. Los cambios detectados pueden ser neutrales en términos de expresión. Es decir, se pueden tener ganancias o pérdidas sin efecto patológico 2. Es poco sensible a los cambios en número de copias 3. Depende del desarrollo de <i>software</i> para su interpretación

Tabla III. Esquema general del desarrollo de biomarcadores tumorales

Etapa	Actividad	Quién la realiza
Fase 0	Desarrollo de la tecnología previa y de los reactivos	Investigadores básicos y empresas de Biotecnología
Fase 1	Desarrollo del ensayo dirigido a los biomarcadores seleccionados	Investigadores aplicados/clínicos
Fase 2	Evaluación inicial de la prevalencia y del uso clínico	Investigadores aplicados/clínicos
Fase 3	Confirmación prospectiva en una serie independiente	Investigadores aplicados/clínicos
Fase 4	Desarrollo comercial	Empresas de Biotecnología

En general, el proceso (que se articula según muestra la **Tabla III**) se puede aplicar a cualquier tipo de biomarcador generado por técnicas analíticas o por técnicas genéticas o genómicas. En esta sección se va a describir el proceso basado en técnicas genómicas, dado que los biomarcadores basados en otro tipo de técnicas analíticas (bioquímica esencialmente) no requieren un tratamiento de los datos tan complejo y tan falto de desarrollo e implementación.

Los *microarrays* de expresión y su empleo para buscar biomarcadores diagnósticos

En general, como se ha comentado anteriormente, el análisis de los patrones de expresión tiene una incidencia directa en el diagnóstico molecular de los tumores. Cuanto mayor sea nuestra capacidad para clasificar un tumor determinado de forma no ambigua, la información

acumulada referente a la biología de ese cáncer (capacidad de metástasis, resistencia al tratamiento, etc.) será más precisa y, en definitiva, mayores serán las posibilidades de éxito en la curación del paciente.

El efecto clínico de los *microarrays* de expresión, aunque esté pendiente de validaciones a gran escala en grupos de pacientes homogéneos, está basado en dos evidencias que ya aparecen documentadas en la literatura. Por un lado, los tumores presentan patrones de expresión directamente relacionados con la línea celular de la que proceden y con características clonales específicas de los pacientes. Por otro lado, los patrones de expresión ayudan a establecer diferencias en la biología y en el pronóstico de los tumores más allá de las aproximaciones empleadas hasta la fecha.

Mediante el microscopio de luz transmitida, los tumores presentan un aspecto que recuerda a la célula normal en la que se origina, pero en la que se hubiera detenido el proceso de diferenciación celular. El resultado de esa similitud es que la taxonomía del cáncer se basa en los distintos tipos histológicos normales en los que se origina. Al aplicar los estudios mediante *arrays* de expresión a diferentes tipos de tumores, lo primero que se observa es que los perfiles de expresión se correlacionan bastante bien con los fenotipos histológicos de los tumores. Como ejemplo, esta relación se ha comprobado y se acerca mucho a una clasificación basada en el linaje celular (12, 13).

Otros autores también han encontrado diferencias en tumores de origen epitelial. Por ejemplo, tumores de mama primarios se agrupan en diferentes patrones de expresión que reflejan el estado de expresión de receptores hormonales y, de alguna manera, los encontrados en líneas celulares derivadas de dos linajes diferentes del tejido mamario (células basales y células lumbinales) (14, 15).

Es muy interesante comprobar que, al margen del tipo celular estudiado o del tipo de análisis, en cualquier *array* la proporción de genes que aparece diferencialmente expresada entre el tumor y el tejido normal de origen es de, aproximadamente, sólo un 5%. Las relaciones que pueden definirse mediante patrones de expresión también pueden contribuir a aumentar el conocimiento de la progresión tumoral. Lesiones precancerosas como las encontradas en adenomas de colon pueden separarse mediante patrones de expresión del tejido normal y del carcinoma. Sin embargo, esos patrones de expresión génica del adenoma presentan un mayor grado de similitud con los carcinomas que con el tejido normal. En general, estos estudios demuestran que el linaje celular es un elemento determinante del patrón de expresión de un tumor.

Otros datos que apoyan el relevante papel del linaje celular provienen de observaciones entre tejidos pareados. Estudios comparativos entre tumores y metástasis o entre momentos terapéuticos distintos que proceden del mismo paciente presentan unos patrones de expresión más similares entre sí que cuando se com-

para el mismo tipo de muestra (tumores, metástasis, después del tratamiento) de diferentes pacientes. Esto se ha comprobado en el cáncer de cabeza y cuello (16), en los melanomas (17) o en el cáncer de mama (18). Dicho compromiso clonal del patrón de expresión de un tumor, íntimamente relacionado con el paciente y su origen de linaje celular, permite obtener marcadores moleculares individuales y, con el tiempo, será muy útil en el desarrollo de terapias individualizadas.

Los patrones de expresión en la búsqueda de biomarcadores de pronóstico y de predicción de la respuesta al tratamiento

Una vez establecido que los patrones de expresión reafirman en gran medida los patrones histológicos diferenciales en los tumores, es necesario demostrar que la ventaja añadida de reconocer esos patrones es poder subclasificar los tumores en función de su comportamiento biológico. Estos subgrupos deberían estar definidos mediante biomarcadores moleculares sutiles y complejos que conlleven un impacto significativo en la predicción del pronóstico y de la respuesta al tratamiento. Dicha aproximación se ha conseguido en varios tipos de cáncer.

Leucemias mieloides agudas (13) y linfomas (19) son tipos de cáncer especialmente informativos por su accesibilidad, la abundancia de marcadores, el conocimiento adquirido de las células normales originales y la existencia de tratamientos muy efectivos. Por ejemplo, también se han publicado análisis de expresión, con *arrays* de Afymetrix, de más de 300 pacientes infantiles con leucemia linfoblástica aguda (LLA) (20). Los resultados permiten clasificar dichas leucemias en subgrupos que relacionan patrones de expresión diferentes con alteraciones citogenéticas y/o histológicas esenciales: LLA de células T, hiperdiploidía con más de 50 cromosomas, BCR-ABL, E2A-PBX, TEL-AML1, reordenamientos del gen MLL, y un grupo diferenciado que carece de marcador citogenético específico. Dado que cada uno de estos subgrupos conlleva un pronóstico asociado, el estudio mediante *microarrays* proporcionará un gran número de marcadores informativos sobre la biología de ese pronóstico y facilitará enormemente la monitorización del proceso terapéutico. En la siguiente década se asistirá a la validación de todos estos marcadores.

Otro ejemplo significativo son los estudios realizados en cáncer de pulmón y próstata, y el establecimiento de grupos de pronóstico (21, 22). Al menos tres estudios han ayudado a separar los adenocarcinomas de pulmón en tres subgrupos que expresan de forma diferencial un conjunto de genes. Dichos subgrupos tienen diferentes patrones de supervivencia: los de peor pronóstico son aquellos adenocarcinomas que expresan de forma preferente genes de diferenciación neumocítica y tienen una alta expresión de genes implicados en procesos de remodelación tisular y actividad proliferativa. Se han publicado estudios de la misma

naturaleza en cáncer de próstata (23). Un cuidadoso análisis que integra datos de cuatro estudios diferentes con *microarrays* ha identificado que las rutas de biosíntesis de las purinas y las poliaminas están completamente desreguladas a nivel transcripcional.

Finalmente, otro de los usos más prometedores de los *microarrays* son los patrones de expresión como predictores de la respuesta al tratamiento. Se han confirmado datos en relación con el tratamiento de linfomas, leucemias agudas y cáncer de mama. El grupo de Alizadeh, en colaboración con el *Lymphoma/Leukemia Molecular Profiling Project*, ha estudiado a 240 pacientes con linfoma B difuso de células grandes mediante *microarrays* de expresión (24, 25). Ha confirmado la existencia de tres subgrupos y ha descrito cinco patrones de expresión diferentes (firmas moleculares) con pronósticos asociados diferenciales. Las firmas del centro germinal, del MHC clase II y del nódulo linfático se asocian a buen pronóstico, mientras que las firmas de genes proliferativos y “otros genes” se asocian a mal pronóstico. Como todos los pacientes recibieron el mismo tratamiento, los resultados de los perfiles de expresión también han podido utilizarse para elaborar predictores de la respuesta al tratamiento estándar de este tipo de linfoma.

Los estudios realizados con muestras de cáncer de mama han sido efectuados con diferentes plataformas y sobre diferentes poblaciones (26, 27). Sin embargo, los resultados son muy concordantes. La primera conclusión es que el elemento de mayor impacto en el perfil de expresión del cáncer de mama es el estatus de receptor de estrógenos (RE), que refuerza el papel fundamental del eje RE en la génesis del cáncer de mama humano y en la determinación de la respuesta a la manipulación hormonal. Además, subdividen los grupos pronósticos con mayor precisión. Sotiriou *et al.* (28)

han llevado a cabo estudios de perfiles asociados a la respuesta al tratamiento y han podido determinar genes directamente relacionados con la respuesta farmacodinámica.

BARRERAS PARA LA IMPLANTACIÓN: INNOVACIÓN CONSTANTE EN ALCANCE Y RESOLUCIÓN, AUSENCIA DE UNA VALIDACIÓN ADECUADA Y DIFICULTADES EN EL ANÁLISIS DE DATOS GENÓMICOS

A pesar de toda la literatura y del conocimiento recogido sobre el tema, lo cierto es que la Genómica y su aplicación clínica son conceptos que sólo ahora comienzan a discurrir de forma cotidiana. Hasta la fecha, los intentos de generar biomarcadores a partir de datos genómicos han sido fundamentalmente infructuosos. Algunos ejemplos excepcionales son el biomarcador basado en la amplificación del gen HER2/NEU en el cáncer de mama (revisada por Pritchard *et al.* [29]) y las plataformas de expresión MammaPrint®, Oncotype DX® (revisadas por Morris y Carey [15]) y H/I® (30), que ya se han empleado en ensayos clínicos en fase II.

Respecto a la amplificación del gen HER2/NEU, no existen dudas (véase Tabla I) de su uso como biomarcador diagnóstico, pronóstico y predictivo, y están recogidas las recomendaciones internacionales para su empleo en los laboratorios de diagnóstico (31).

En cuanto a los perfiles de expresión, la situación está menos definida. En la actualidad, se han aprobado para su uso clínico tres ensayos o test de pronóstico para cáncer de mama que se basan estrictamente en perfiles de expresión génica (recogidos con sus características resumidas en la **Tabla IV**).

Tabla IV. Resumen de las plataformas genómicas aceptadas clínicamente como biomarcadores para el estudio del cáncer de mama

Nombre	Comercializado	Material biológico	Número de genes	Tecnología empleada	Tipo de biomarcador	Empleado para
MammaPrint®	Agendia BV (Amsterdam, Países Bajos)	ARN	70	<i>Microarray</i>	Pronóstico y predictivo (¿?)	Asignar a una paciente a un grupo de riesgo y calcular su posibilidad de recurrencia
Oncotype DX®	Genomic Health (Redwood City, California)	Biopsia incluida en parafina	21	RT-PCR	Pronóstico y predictivo	Cáncer invasivo en estadio temprano, positivo para receptor de estrógenos y negativo para ganglios
H/I®	AvariaDX (Carlsbad, California)	ARN	2	RT-PCR	Pronóstico	Asignar a una paciente a un grupo de riesgo

Recientemente, se han publicado un par de comparaciones entre las tres plataformas ya disponibles para ensayos de fase II (30, 32). Las conclusiones de ambos estudios son semejantes: Oncotype DX® es una herramienta sólida para predecir la aparición de metástasis y la quimioterapia es más efectiva si se asigna según el resultado de este test (siempre teniendo en cuenta que es un grupo muy bien definido de pacientes). La evidencia de la utilidad clínica del test MammaPrint® no es tan clara y la capacidad para predecir el efecto de la quimioterapia no está demostrada. Finalmente, el test H/I® es el menos desarrollado. En general, las conclusiones del estudio comparativo reflejan que son necesarias más observaciones clínicas para asignar a estos biomarcadores su verdadero papel.

Existen en el mercado otros test genéticos de biomarcadores basados en la presencia de mutaciones de único gen con poder diagnóstico, predictivo y pronóstico, generalmente relacionados con una terapia molecular dirigida a paliar los efectos de esas mutaciones en cada gen afectado. Este tipo de marcadores y las terapias asociadas se detallan y discuten en otros capítulos de este documento.

Esta situación descompensada (miles de artículos frente a menos de diez test genómicos disponibles como biomarcadores en la práctica clínica) tiene su origen en las numerosas dificultades que aparecen en el desarrollo de biomarcadores en cada una de las fases recogidas en la Tabla III. A continuación, se indican de forma resumida las más relevantes.

Fase 0. Desarrollo de la tecnología previa y de los reactivos

En esta fase las dificultades provienen de la propia naturaleza de la investigación biotecnológica. Cada día aparecen plataformas genómicas de mayor calidad y de mayor resolución. Esto implica que las herramientas asociadas tanto para realizar el ensayo como para obtener resultados han de mejorarse constantemente y la transferencia de la rutina del diagnóstico (con protocolos sencillos, normalizados y casi automáticos) está prácticamente condenada al fracaso.

Fase 1. Desarrollo del ensayo dirigido a los biomarcadores seleccionados

En gran medida las dificultades provienen de la complejidad del análisis de datos. Estudios genómicos masivos, como los realizados con *microarrays* de expresión, ofrecen información rápida y reproducible sobre el nivel de expresión de un número elevado de genes (generalmente, varios miles) en la muestra estudiada en un solo experimento. Dicha información tiene varios niveles de complejidad. El nivel más bajo corresponde a datos simples sobre genes concretos. Un *array* de expresión puede identificar el nivel de

expresión de un gen individual asociado a un determinado fenotipo de forma similar a como ocurre en experimentos más tradicionales. Los niveles más altos de complejidad, sin embargo, permiten examinar la figura completa del conjunto del transcriptoma, es decir, todos los genes que se expresan de forma asociada a ese fenotipo.

Los *arrays* de expresión han acelerado en varios órdenes de magnitud el descubrimiento de marcadores moleculares de posible utilidad en la Medicina Clínica. Las investigaciones recientes procedentes de varios laboratorios están revelando principios fundamentales que ayudan a clasificar una enfermedad de acuerdo con su perfil de expresión específico. Esos principios son sorprendentemente reproducibles, a menudo esclarecedores y algunas veces inesperados. En cualquier caso, sólo estudios multicéntricos (con una muestra de tamaño adecuado) y desarrollados con un gran soporte bioinformático son capaces de producir una selección de biomarcadores sólida. **Éste es el paso en el que nos encontramos en la gran mayoría de los tumores.**

En este punto es esencial recordar que gran parte de las dificultades para avanzar en este punto tienen su origen en la complejidad del análisis de datos. Las herramientas bioinformáticas de análisis requieren un nivel muy alto de experiencia previa, sólo disponible en grandes grupos de investigación. Es el cuello de botella de todo el proceso.

Fase 2. Evaluación de la prevalencia

Esta fase es muy dependiente de la fase anterior y, por ahora, sólo se ha desarrollado para los test mencionados anteriormente (MammaPrint®, Oncotype DX® y H/I®). Para el resto de los tumores no hay datos contrastables.

Fase 3. Confirmación prospectiva y validación en una serie independiente

Aquí la situación está todavía más atrasada, ya que no existe un acuerdo claro sobre la metodología de validación aplicable a cada tipo de plataforma.

En la **Tabla V** se recogen de forma esquemática las dificultades detectadas en cada una de las fases de desarrollo de biomarcadores. Dado que el proceso del desarrollo de biomarcadores requiere la conclusión de una fase para acceder a la siguiente, las dificultades descritas detienen de forma efectiva el proceso en los estadios iniciales.

CONCLUSIONES Y PERSPECTIVAS

La comunidad científica y sanitaria española ya puede acceder a los análisis de expresión mediante *microarrays* de ADNc o de oligos. Los resultados, además, fascinan a los investigadores, puesto que son bastante re-

Tabla V. Resumen de las dificultades y sus consecuencias para desarrollar biomarcadores de origen genómico

Actividad	Dificultades detectadas	Consecuencias
Desarrollo de la tecnología previa y de los reactivos	<ul style="list-style-type: none"> • Innovación constante de la tecnología 	<ul style="list-style-type: none"> • Impide realizar ensayos de reproducibilidad • Compromete el análisis de sesgos tecnológicos • Compromete la validación en series independientes • Modifica el diseño experimental previo
Desarrollo del ensayo de biomarcadores seleccionados	<ul style="list-style-type: none"> • Competencia de tecnologías similares • Ausencia de desarrollos bioinformáticos comunes y aceptados 	<ul style="list-style-type: none"> • Análisis de datos adaptados al experimento • Ausencia de consenso sobre la metodología estadística y de análisis • Impide la validación en series independientes y la traslación a la clínica
Evaluación inicial de la prevalencia y del uso clínico	<ul style="list-style-type: none"> • Financiación insuficiente • Series de muestras de número limitado 	<ul style="list-style-type: none"> • Estimación errónea de la incidencia real • Estimación errónea para atribuir el carácter predictivo o pronóstico del marcador
Confirmación prospectiva en una serie independiente	<ul style="list-style-type: none"> • Financiación insuficiente • Series de muestras de número limitado 	<ul style="list-style-type: none"> • Ausencia de validación • Dificultades para la aceptación del biomarcador por las agencias reguladoras
Desarrollo comercial	<ul style="list-style-type: none"> • Logística del análisis: implementación, tecnología, certificación, etc. 	<ul style="list-style-type: none"> • Precio elevado del diagnóstico • Falta de rigor en los controles • Adaptaciones del análisis no validadas por la comunidad científica

producibles y aportan una gran cantidad de información sobre la regulación de la expresión génica en condiciones normales y patológicas. Como consecuencia del éxito de estas herramientas, se recogen en la literatura un gran número de experimentos realizados con *arrays* de expresión y que agrupan, relacionan o describen los perfiles de expresión de cientos o miles de genes según dos tipos generales de preguntas:

- ¿Se pueden utilizar los perfiles de expresión para clasificar mejor los diferentes tipos de cáncer?
- ¿Se puede conocer el efecto en el perfil de expresión de estímulos extrínsecos (drogas, estrés, etc.) o internos (mutaciones, sobreexpresión de genes)?

La respuesta a ambas preguntas es afirmativa, aunque, en general, este tipo de análisis han de ser validados con posterioridad mediante técnicas de PCR cuantitativa. Se debe a que existen problemas de estandarización y otras fuentes de variabilidad como: 1) naturaleza normalmente lábil y diluida de las muestras biológicas; 2) condiciones estocásticas y subóptimas de los procesos de marcaje y de interacción molecular con elementos microscópicos insuficientemente redundantes; 3) sensibilidad indeterminada a factores y procedimientos mal estandarizados y controlados (temperatura y volúmenes de las disoluciones, concentración de ozono atmosférico); 4) dependencia en instrumentación calibrada de forma variable (baños ter-

mostatizados, escáneres); 5) diversidad de los procedimientos de procesamiento de datos; 6) sistemas inapropiados de análisis de datos; y 7) escasez de recursos bioinformáticos para el tratamiento y comprensión de datos masivos. A todo esto se añade la diversidad existente de alternativas y de plataformas tecnológicas, y la variabilidad del proceso de producción de plataformas “hechas en casa” por laboratorios de apoyo a la investigación de centros académicos.

Con todo ello, el mensaje es que la tecnología de origen genómico está produciendo herramientas de uso clínico (biomarcadores) que van a contribuir de forma esencial a mejorar los diagnósticos y la efectividad de los tratamientos. De momento, estas aplicaciones están limitadas al cáncer de mama, aunque se adivinan procesos en otros tipos tumorales. Las dificultades para implantar de manera más rápida y generalizada este tipo de tecnología radica en gran medida en la complejidad del análisis de datos, una actividad científica de difícil acceso para la comunidad clínica y científica en general.

En resumen, esta tecnología es revolucionaria y va a tener un gran impacto en la Medicina. Sin embargo, ninguna tecnología es suficiente para contribuir por sí sola al avance de las grandes cuestiones abiertas en la investigación biosanitaria. Lo importante son las preguntas y establecer validaciones adecuadas que permitan la traducción de todos estos hallazgos a la práctica clínica de rutina.

BIBLIOGRAFÍA

1. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. *Nat Med* 2004; 10 (8): 789-99.
2. Cigudosa JC. The microarray revolution in biomedical research: types of platforms, uses and perspectives in oncology. *An Sist Sanit Navar* 2004; 27 (1): 11-20.
3. Miller LD, *et al.* Optimal gene expression analysis by microarrays. *Cancer Cell* 2002; 2 (5): 353-61.
4. Kallioniemi A, *et al.* Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 1992; 258 (5083): 818-21.
5. Fiegler H, *et al.* DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones. *Genes Chromosomes Cancer* 2003; 36 (4): 361-74.
6. Ishkanian AS, *et al.* A tiling resolution DNA microarray with complete coverage of the human genome. *Nat Genet* 2004; 36 (3): 299-303.
7. Solinas-Toldo S, *et al.* Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* 1997; 20 (4): 399-407.
8. Pollack JR, *et al.* Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet* 1999; 23 (1): 41-6.
9. Lucito R, *et al.* Detecting gene copy number fluctuations in tumor cells by microarray analysis of genomic representations. *Genome Res* 2000; 10 (11): 1726-36.
10. Mei R, *et al.* Genome-wide detection of allelic imbalance using human SNPs and high-density DNA arrays. *Genome Res* 2000; 10 (8): 1126-37.
11. Kaller M, Lundeberg J, Ahmadian A. Arrayed identification of DNA signatures. *Expert Rev Mol Diagn* 2007; 7 (1): 65-76.
12. Bullinger L, *et al.* Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N Engl J Med* 2004; 350 (16): 1605-16.
13. Bullinger L, Valk PJ. Gene expression profiling in acute myeloid leukemia. *J Clin Oncol* 2005; 23 (26): 6296-305.
14. Cheang MC, van de Rijn M, Nielsen TO. Gene expression profiling of breast cancer. *Annu Rev Pathol* 2008; 3: 67-97.
15. Morris SR, Carey LA. Gene expression profiling in breast cancer. *Curr Opin Oncol* 2007; 19 (6): 547-51.
16. Chen ZG. Exploration of metastasis-related proteins as biomarkers and therapeutic targets in the treatment of head and neck cancer. *Curr Cancer Drug Targets* 2007; 7 (7): 613-22.
17. Riker AI, *et al.* The gene expression profiles of primary and metastatic melanoma yields a transition point of tumor progression and metastasis. *BMC Med Genomics* 2008; 1 (1): 13.
18. Nuyten DS, van de Vijver MJ. Using microarray analysis as a prognostic and predictive tool in oncology: focus on breast cancer and normal tissue toxicity. *Semin Radiat Oncol* 2008; 18 (2): 105-14.
19. Leich E, *et al.* Diagnostic and prognostic significance of gene expression profiling in lymphomas. *Apmis* 2007; 115 (10): 1135-46.
20. Pui CH, Evans WE. Treatment of acute lymphoblastic leukemia. *N Engl J Med* 2006; 354 (2): 166-78.
21. Garber ME, *et al.* Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci USA* 2001; 98 (24): 13784-9.
22. Virtanen C, *et al.* Integrated classification of lung tumors and cell lines by expression profiling. *Proc Natl Acad Sci USA* 2002; 99 (19): 12357-62.
23. Singh D, *et al.* Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* 2002; 1 (2): 203-9.
24. Rosenwald A, *et al.* The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* 2002; 346 (25): 1937-47.
25. Rosenwald A, *et al.* The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma. *Cancer Cell* 2003; 3 (2): 185-97.
26. Van de Vijver MJ, *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002; 347 (25): 1999-2009.
27. Van 't Veer LJ, *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002; 415 (6871): 530-6.
28. Sotiriou C, *et al.* Gene expression profiles derived from fine needle aspiration correlate with response to systemic chemotherapy in breast cancer. *Breast Cancer Res* 2002; 4 (3): R3.
29. Pritchard KI, *et al.* HER-2 and topoisomerase II as predictors of response to chemotherapy. *J Clin Oncol* 2008; 26 (5): 736-44.
30. Marchionni L, *et al.* Impact of gene expression profiling tests on breast cancer outcomes. *Evid Rep Technol Assess (Full Rep)* 2007; 160: 1-105.
31. Wolff AC, *et al.* American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *J Clin Oncol* 2007; 25 (1): 118-45.
32. Marchionni L, *et al.* Systematic review: gene expression profiling assays in early-stage breast cancer. *Ann Intern Med* 2008; 148 (5): 358-69.